

military conquest. The language triumphed because its speakers became more numerous and prosperous by growing their own food instead of relying on hunting.

Regardless of how Indo-European diffused, communication was poor among different peoples, whether warriors or farmers. After many generations of complete isolation, individual groups evolved increasingly distinct languages.

### KEY ISSUE 3

## Where Are Other Language Families Distributed?

- Classification of Languages
- Distribution of Language Families

This section describes where different languages are found around the world. The several thousand spoken languages can be organized logically into a small number of language families. Larger language families can be further divided into language branches and language groups. ■

## Classification of Languages

Figure 5-16 shows the world's language families:

- A language in the *Indo-European family*, such as English, is spoken by 46 percent of the world's people.
- A language in the *Sino-Tibetan family*, such as Mandarin, is spoken by 21 percent of the world, mostly in China.
- A language in the *Afro-Asiatic family*, including Arabic, is spoken by 6 percent, mostly in the Middle East
- A language in the *Austronesian family* is spoken by 6 percent, mostly in Southeast Asia.
- A language in the *Niger-Congo family* is spoken by 6 percent, mostly in Africa.
- A language in the *Dravidian family* is spoken by 4 percent, mostly in India.
- A language in the *Altaic family* is spoken by 2 percent, mostly in Asia.
- A language in the *Austro-Asiatic family* is spoken by 2 percent, mostly in Southeast Asia.
- *Japanese*, a separate language family, is spoken by 2 percent.
- The remaining 5 percent of the world's people speak a language belonging to one of 100 smaller families.

Figure 5-17 attempts to depict differences among language families, branches, and groups. Language families form the trunks of the trees, whereas individual languages are displayed as leaves. The larger the trunks and leaves are, the greater the number of speakers of those families and languages. Some trunks divide into several branches, which logically represent

language branches. The branches representing Germanic, Balto-Slavic, and Indo-Iranian in Figure 5-17 divide a second time into language groups.

Figure 5-17 displays each language family as a separate tree at ground level because differences among families predate recorded history. Linguists speculate that language families were joined together as a handful of superfamilies tens of thousands of years ago. Superfamilies are shown as roots below the surface because their existence is highly controversial and speculative.

## Distribution of Language Families

Nearly one-half the people in the world speak an Indo-European language. The second-largest family is Sino-Tibetan, spoken by one-fifth of the world. Another half-dozen families account for most of the remainder.

### Sino-Tibetan Family

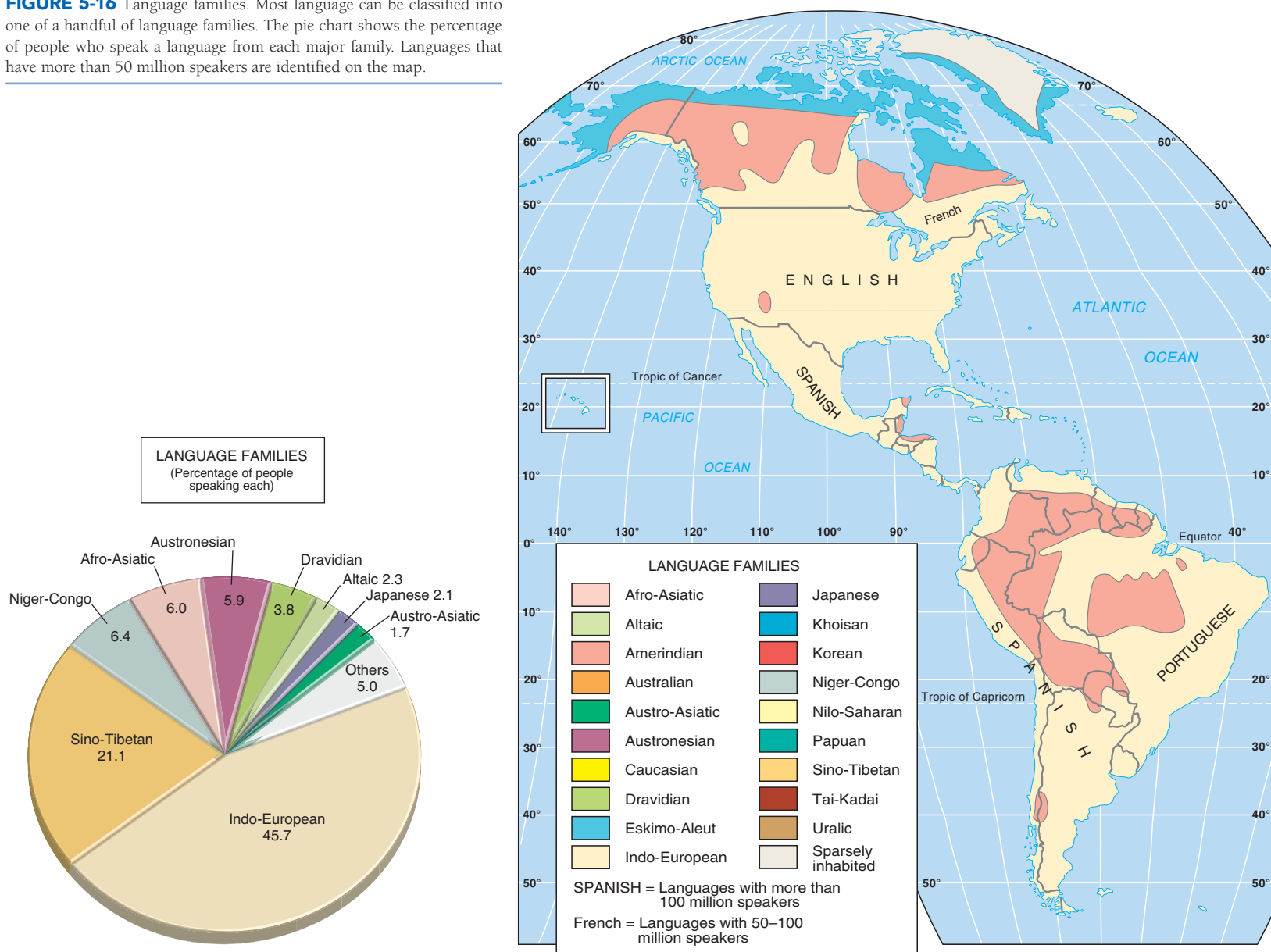
The Sino-Tibetan family encompasses languages spoken in the People's Republic of China—the world's most populous state at more than 1 billion—as well as several smaller countries in Southeast Asia. The languages of China generally belong to the Sinitic branch of the Sino-Tibetan family.

There is no single Chinese language. Rather, the most important is Mandarin (or, as the Chinese call it, *pu tong hua*—“common speech”). Spoken by approximately three-fourths of the Chinese people, Mandarin is by a wide margin the most used language in the world. Once the language of emperors in Beijing, Mandarin is now the official language of both the People's Republic of China and Taiwan, as well as one of the six official languages of the United Nations. Other Sinitic branch languages are spoken by tens of millions of people in China, mostly in the southern and eastern parts of the country—Wu, Yue (also known as Cantonese), Min, Jinyu, Xiang, Hakka, and Gan. However, the Chinese government is imposing Mandarin countrywide.

The relatively small number of languages in China (compared to India, for example) is a source of national strength and unity. Unity is also fostered by a consistent written form for all Chinese languages. Although the words are pronounced differently in each language, they are written the same way.

You already know the general structure of Indo-European quite well because you are a fluent speaker of at least one Indo-European language. But the structure of Chinese languages is quite different (Figure 5-18). They are based on 420 one-syllable words. This number far exceeds the possible one-syllable sounds that humans can make, so Chinese languages use each sound to denote more than one thing. The sound *shi*, for example, may mean “lion,” “corpse,” “house,” “poetry,” “ten,” “swear,” or “die.” The sound *jian* has more than 20 meanings, including “to see.” The listener must infer the meaning from the context in the sentence and the intonation the speaker uses. In addition, two one-syllable words can be combined into two syllables, forming a new word. For example, the two-syllable word “Shanghai” is a

**FIGURE 5-16** Language families. Most language can be classified into one of a handful of language families. The pie chart shows the percentage of people who speak a language from each major family. Languages that have more than 50 million speakers are identified on the map.



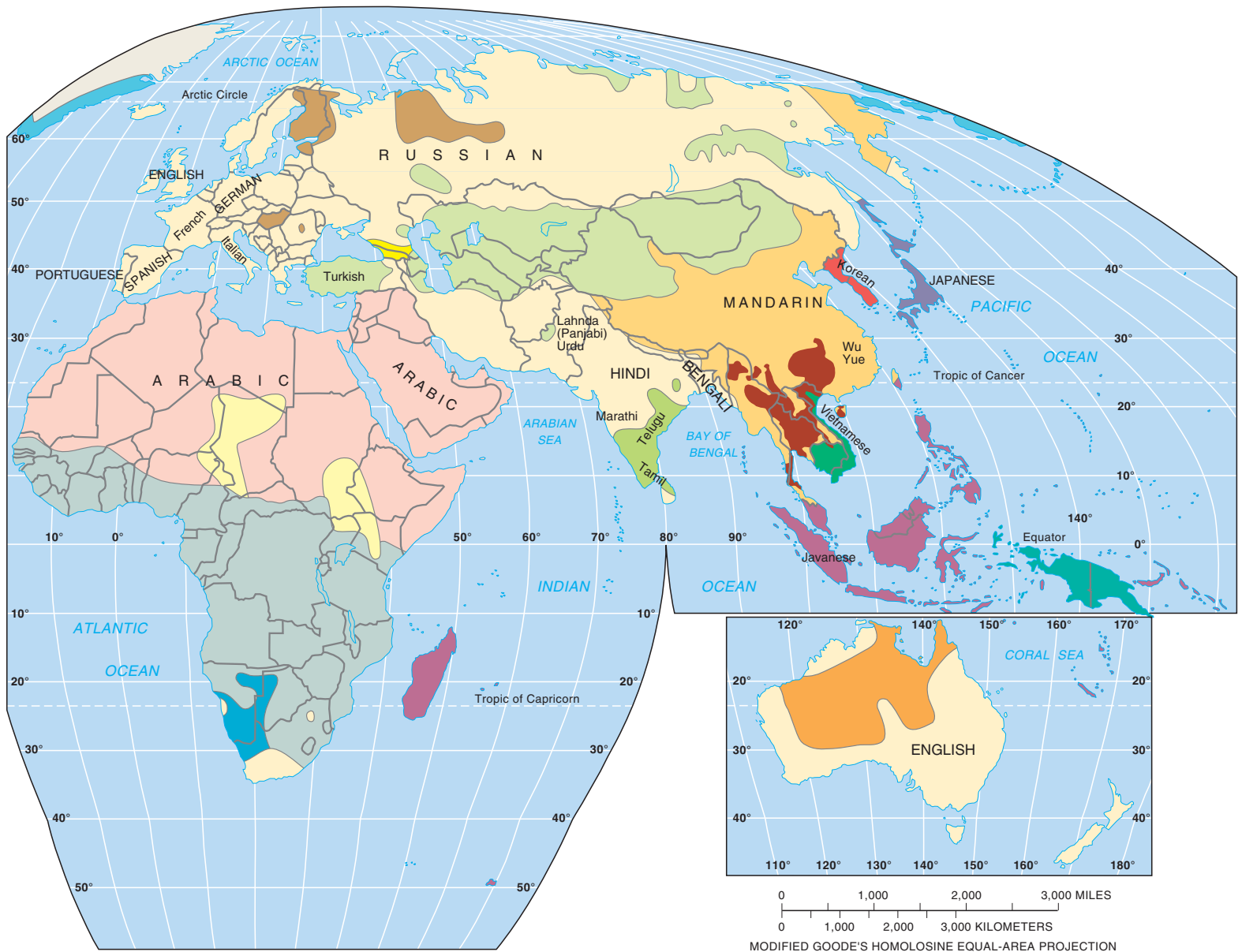
combination of words that mean “above” and “sea.” *Kan jian*—a combination of the words for “look” and “see,” which would be redundant in English—clarifies that “to see” is the intended meaning for the multiple meanings of *jian*.

The other distinctive characteristic of the Chinese languages is the method of writing. The Chinese languages are written with a collection of thousands of characters. Some of the characters represent sounds pronounced in speaking, as in English. However, most are **ideograms**, which represent ideas or concepts, not specific pronunciations. The system is intricate and mature, having developed over 4,000 years. The main language problem for the Chinese is the difficulty in learning to write because of the large number of characters. The Chinese government reports that 16 percent of the population over age 16 is unable to read or write more than a few characters.

### Other East and Southeast Asian Language Families

In addition to Sino-Tibetan, several other language families spoken by large numbers of people can be found in East and Southeast Asia. If you look at their distribution in Figure 5-16, you can see a physical reason for their independent development: These language families are clustered either on islands or peninsulas.

- Austronesian.** Spoken by about 6 percent of the world’s people, speakers of Austronesian languages are mostly in Indonesia, the world’s fourth most populous country. With its inhabitants dispersed among thousands of islands, Indonesia has an extremely large number of distinct



languages and dialects; 722 actively used languages are identified by *Ethnologue*. Indonesia's most widely used first language is Javanese, spoken by 85 million people, mostly on the island of Java, where two-thirds of the country's population is clustered.

Language maps show a striking oddity: The people of Madagascar, the large island off the east coast of Africa, speak Malagasy, which belongs to the Austronesian family, even though the island is 3,000 kilometers (1,900 miles) distant from any other Austronesian-speaking country. This is strong evidence of migration to Madagascar from present-day Indonesia. Malayo-Polynesian people apparently sailed in small boats across the Indian Ocean to reach Madagascar approximately 2,000 years ago.

- **Austro-Asiatic.** Spoken by about 2 percent of the world's population, Austro-Asiatic is based in Southeast Asia. Vietnamese, the most spoken tongue of the Austro-Asiatic language family, is written with our familiar Roman alphabet, with the addition of a large number of diacritical marks above the vowels. The Vietnamese alphabet was devised in the seventh century by Roman Catholic missionaries.
- **Tai Kadai.** Once classified as a branch of Sino-Tibetan, the principal languages of this family are spoken in Thailand and neighboring portions of China. Similarities with the Austronesian family lead some linguistic scholars to speculate that people speaking these languages may have migrated from the Philippines.
- **Japanese.** Written in part with Chinese ideograms, Japanese also uses two systems of phonetic symbols, like

Western languages, used either in place of the ideograms or alongside them. Chinese cultural traits have diffused into Japanese society, including the original form of writing the Japanese language. But the structures of the two languages differ. Foreign terms may be written with one of these sets of phonetic symbols.

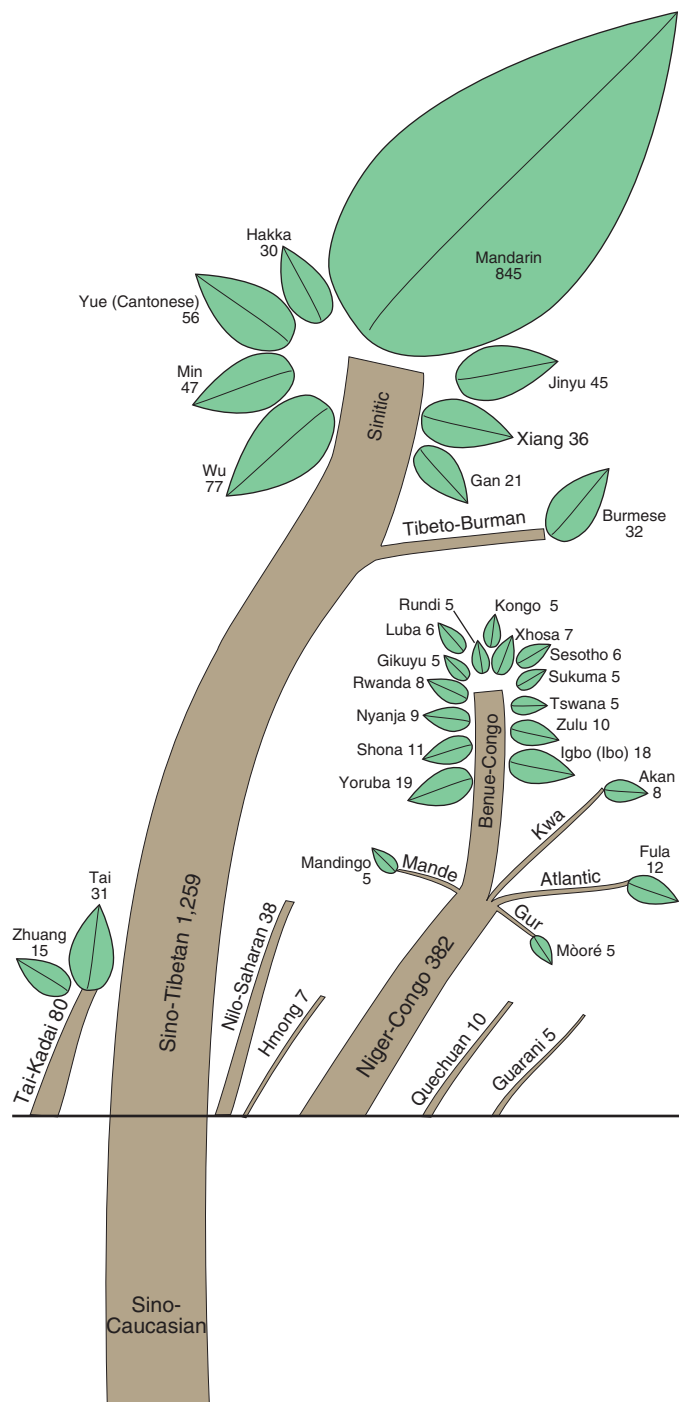
- **Korean.** Usually classified as a separate language family, Korean may be related to the Altaic languages of Central Asia or to Japanese. Unlike Sino-Tibetan languages and Japanese, Korean is written not with ideograms but in a system known as *hankul* (also called *hangul* and *onmun*). In this system, each letter represents a sound, as in Western languages. More than half of the Korean vocabulary derives from Chinese words. In fact, Chinese and Japanese words are the principal sources for creating new words to describe new technology and concepts.

## Languages of the Middle East and Central Asia

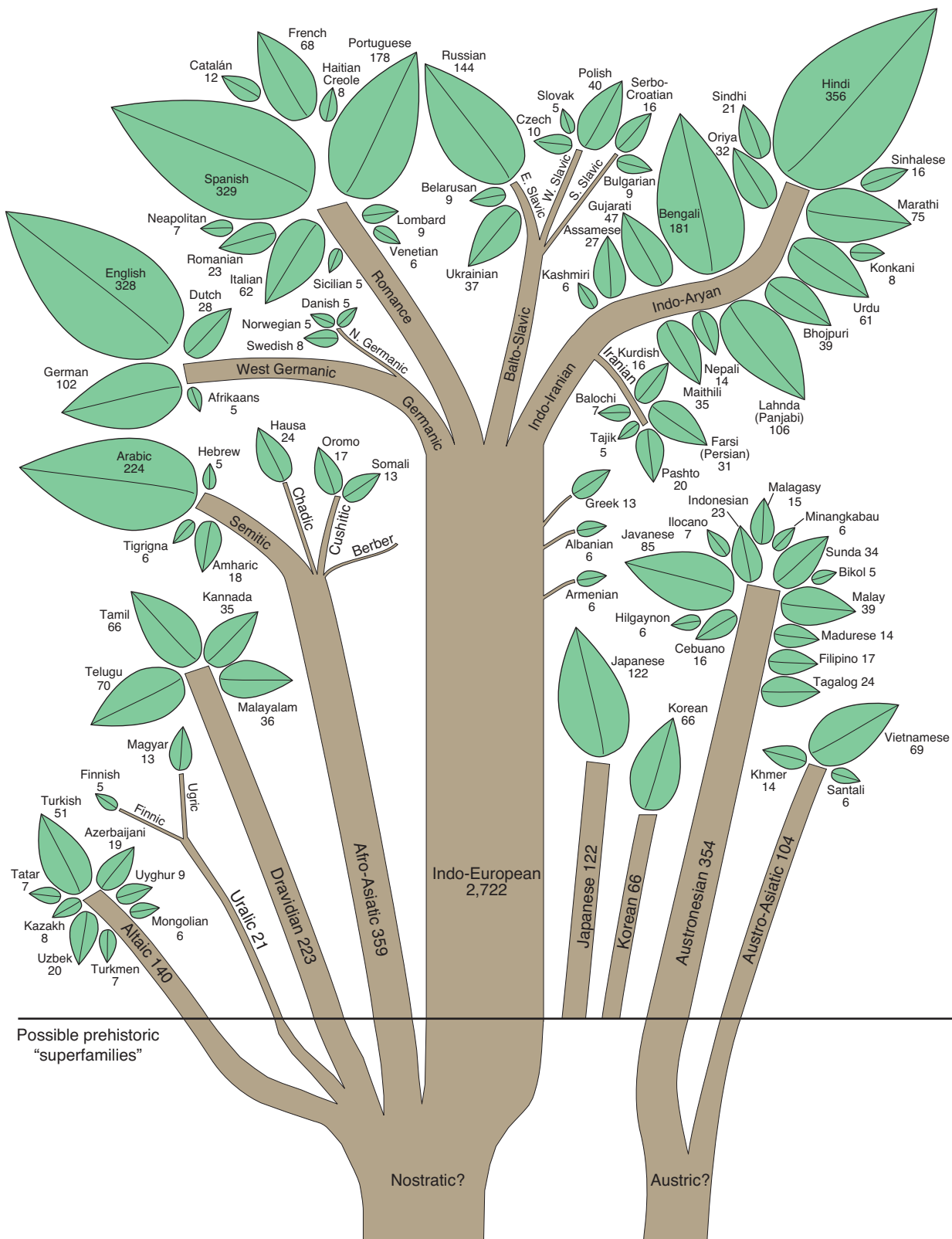
Major language families in the Middle East and Central Asia include Afro-Asiatic and Altaic. Uralic languages were once classified with Altaic.

- **Afro-Asiatic.** Arabic is the major language of this family, an official language in two dozen countries of the Middle East, and one of six official languages of the United Nations. In addition to the 200-million-plus native speakers of Arabic, a large percentage of the world's Muslims have at least some knowledge of Arabic because Islam's holiest book, the Quran (Koran), was written in that language in the seventh century. This family also includes Hebrew, the language of the Judeo-Christian Bible.
- **Altaic.** These languages are thought to have originated in the steppes bordering the Qilian Shan and Altai mountains between Tibet and China. Present distribution covers an 8,000-kilometer (5,000-mile) band of Asia. The Altaic language with by far the most speakers is Turkish. Turkish was once written with Arabic letters. But in 1928 the Turkish government, led by Kemal Ataturk, ordered that the language be written with the Roman alphabet instead. Ataturk believed that switching to Roman letters would help modernize the economy and culture of Turkey through increased communications with European countries.

When the Soviet Union governed most of the Altaic-speaking region of Central Asia, use of Altaic languages was suppressed to create a homogeneous national culture. One element of Soviet policy was to force everyone to write with the Russian Cyrillic alphabet, even though some had traditionally employed Arabic letters. With the dissolution of the Soviet Union in the early 1990s, Altaic languages became official in several newly independent countries, including Azerbaijan, Kazakhstan, Kyrgyzstan, Turkmenistan, and Uzbekistan. People in these countries are no longer forced to learn Russian and write with Cyrillic letters.



**FIGURE 5-17** Language family tree. Language families are divided into branches and groups. Shown here are language families and individual languages that have more than 5 million speakers. Numbers on the tree are in millions of speakers. Below ground level, the language tree's "roots" are shown. However, the theory that several language families had common origins tens of thousands of years ago is a highly controversial speculation advocated by some linguists and rejected by others.



- **Uralic.** Every European country is dominated by Indo-European speakers, except for three—Estonia, Finland, and Hungary (refer to Figure 5-9). The Estonians, Finns, and Hungarians speak languages that belong to the Uralic family.

The Altaic and Uralic language families were once thought to be linked as one family, but recent studies point to geographically distinct origins. Uralic languages are traceable back to a common language, Proto-Uralic, first used 7,000 years ago by people living in the Ural Mountains



**FIGURE 5-18** Chinese language ideograms. The Chinese languages are written with ideograms, most of which represent ideas or concepts rather than sounds.

of present-day Russia, north of the Kurgan homeland. Migrants carried the Uralic languages to Europe, carving out homelands for themselves in the midst of Germanic- and Slavic-speaking peoples and retaining their language as a major element of cultural identity.

## African Language Families

No one knows the precise number of languages spoken in Africa, and scholars disagree on classifying those known into families. In the 1800s, European missionaries and colonial officers began to record African languages using the Roman or Arabic alphabet. More than 1,000 distinct languages and several thousand named dialects have been documented. Most lack a written tradition.

Figure 5-19 shows the broad view of African language families, and Figure 5-20 hints at the complex pattern of multiple tongues of Nigeria. This great number of languages results from at least 5,000 years of minimal interaction among the thousands of cultural groups inhabiting the African continent. Each group developed its own language, religion, and other cultural traditions in isolation from other groups.

In northern Africa the language pattern is relatively clear, because Arabic, an Afro-Asiatic language, dominates, although in a variety of dialects. In sub-Saharan Africa, however, languages grow far more complex.

- **Niger-Congo.** More than 95 percent of the people in sub-Saharan Africa speak languages of the Niger-Congo family. One of these languages—Swahili—is the first language of

only 800,000 people and an official language in only one country (Tanzania), but it is spoken as a second language by approximately 30 million Africans.

Especially in rural areas, the local language is used to communicate with others from the same village, and Swahili is used to communicate with outsiders. Swahili originally developed through interaction among African groups and Arab traders, so its vocabulary has strong Arabic influences. Also, Swahili is one of the few African languages with an extensive literature.

- **Nilo-Saharan.** Languages of this family are spoken by a few million people in north-central Africa, immediately north of the Niger-Congo language region. Divisions within the Nilo-Saharan family exemplify the problem of classifying African languages. Despite fewer speakers, the Nilo-Saharan family is divided into six branches, plus numerous groups and subgroups. The total number of speakers of each individual Nilo-Saharan language is extremely small.
- **Khoisan.** A distinctive characteristic of the Khoisan languages is the use of clicking sounds. Upon hearing this, whites in southern Africa derisively and onomatopoeically named the most important Khoisan language Hottentot.

## KEY ISSUE 4

# Why Do People Preserve Local Languages?

- Preserving Language Diversity
- Global Dominance of English

The distribution of a language is a measure of the fate of an ethnic group. English has diffused around the world from a small island in northwestern Europe because of the cultural dominance of England and the United States over other territory on Earth's surface. Icelandic remains a little-used language because of the isolation of the Icelandic people.

As in other cultural traits, language displays the two competing geographic trends of globalization and local diversity. English has become the principal language of communication and interaction for the entire world. At the same time, local languages endangered by the global dominance of English are being protected and preserved. ■

## Preserving Language Diversity

Thousands of languages are **extinct languages** once in use—even in the recent past—but no longer spoken or read in daily activities by anyone in the world. *Ethnologue* considers 473 languages as nearly extinct because only a few older speakers are